ORC 66-43
DECEMBER 1966

A ~~SINGLE~~ Simple PROOF OF: $L=\lambda W$

by

William S. Jewell

OPERATIONS RESEARCH CENTER

COLLEGE OF ENGINEERING

UNIVERSITY OF CALIFORNIA - BERKELEY

A SIMPLE PROOF OF: $L = \lambda W$

by

William S. Jewell
Operations Research Center
University of California
Berkeley, California

December 1966 ORC 66-43

This research has been partially supported by the Office of Naval Research under Contract Nonr-222(83), the National Science Foundation under Grant GP-4593, and the U.S. Army Research Office-Durham, under Contract DA-31-124-ARO-D-331 with the University of California.

ABSTRACT

A simple proof of the fundamental queueing formula $L=\lambda W$ is given which is based on renewal theory. The basic assumptions which are needed are: (1) the event {system is empty} is recurrent, and (2) the arrival and waiting-time mechanisms are reset by the next arrival after this event occurs.

A SIMPLE PROOF OF: $L = \lambda W$

by

W. S. Jewell
Operations Research Center
Berkeley, California

J.D.C. Little's proof [3] of "$L = \lambda W$" ranks as one of the most important unifying results of queueing theory. However, as Little himself has remarked, in a private communication: "the author must be congratulated for the rigor of his presentation, but he might have explained the ideas a little more".

The following proof has the advantage that it relies only on renewal theory; the somewhat stronger assumptions which are needed are directly related to usual queueing concepts and are, moreover, satisfied in most congestion models. In this way, the construction of the proof reveals the essential simplicity of the result.

NOTATION

Units *arrive* at, *wait* in, and then *leave* from some well-defined *queueing system*. The basic (nonnegative) queueing random variables are:

(1)

$\eta(t)$ -- number of units present in the system at time t. $(-\infty<t<+\infty)$

τ_i -- interval between the epochs of arrival of the $(i-1)^{st}$ and i^{th} units $(-\infty<i<+\infty)$

ω_i -- wait in the system of the i^{th} unit $(-\infty<i<\infty)$

We assume any interaction of these variables is *temporally homogeneous* i.e., the selection of a time origin is arbitrary. We then select this origin and number the units so that the *zero*th unit arrives at time $t = 0$.

Assuming that $\eta(t)$ will reach zero, with probability one, at some future epoch, we define the familiar related non-negative random variables:

$$\nu = \min\left(0 < n \leq \infty \,\middle|\, \eta\left(\sum_{i=1}^{n} \tau_i^-\right) = 0\right)$$ -- *the number of units processed* (arrived, waited, and left) during the first busy cycle (in addition to those initially present).

(2) $$\gamma = \sum_{i=1}^{\nu} \tau_i$$ -- duration of first *busy cycle.*

β = epoch of last departure before time γ -- duration of first *busy period.*

$\iota = \gamma - \beta$ -- duration of *idle period* prior to arrival of ν^{th} unit (which starts new busy cycle).

Smaller definitions apply to successive busy periods, for which $\eta(0^-) = 0$.

ASSUMPTIONS AND RESULTS

The first basic Assumption is:

Assumption I: *The event* $\mathcal{E} = \{\eta(t) = 0\}$ *, is a* recurrent event *for any*

(3) *given initial condition of the system.*

Thus, for any current condition of the system, $\eta(t)$ will become zero, with probability one, at some future epoch.

Theorem I.

For any *realization of the random variables (1) and (2) , with* $\eta(0-) = 0$ *and the zero*th *unit arriving at time zero:*

(4) $$\sum_{i=0}^{\nu-1} \omega_i = \int_0^t \eta(u)\,du \qquad \beta \leq t < \gamma ,$$

when finite.

Proof:

Figure 1 shows a typical realization of a busy cycle, with $\eta(0-) = 0$ and ν , γ finite. The upper curve is the cumulative number of arrivals in $[0,t]$, and the lower curve shows the cumulative number of departures in the same interval (which is defined by the upper curve, the $\{\omega_i\}$, and the internal mechanism of the system which rearranges order of departure). Since $\eta(t)$ is the difference in ordinate between these two curves, by definition, the RHS in (4) is just the shaded area shown for all $t \in [\beta,\gamma)$.

Any unit is waiting in the system at time t if and only if its epoch of arrival is $\leq t$, and its epoch of departure is $> t$ [3]. In Figure 1, we have indicated the values of $\omega_0,\omega_1, \ldots, \omega_{\nu-1}$ *assuming departure in order of arrival:* however, this assumption is not needed to note that the shaded area is also just the sum on the LHS of (4), no matter how the waiting times of units in the system are rearranged, since each jump in $\eta(t)$ has unity magnitude.

For an *arbitrary initial busy cycle* in which $\eta(0-) = n_0 > 0$, (4) is obviously still correct if we add to the sum on the LHS the *residual* waiting times of the items present at time zero.

We now require the following rather complicated "reset" Assumption:

Assumption II: *Whenever* $\eta(t)$ *reaches zero, the arrival and waiting-time mechanisms are "reset" by the next arrival, i.e. the joint distribution of*

(5) $\{\nu;\tau_1,\tau_2, \ldots, \tau_\nu ; \omega_0,\omega_1, \ldots, \omega_{\nu-1}, \iota\}$ *is identical for each busy cycle and independent from one busy cycle to the next (units renumbered for each cycle as in Figure 1).*

We also henceforth exclude the trivial possibility that $E\{\gamma\} = 0$, or $E(\nu) = 0$, and interpret $(1/\infty) = 0$, and $(\infty/1) = \infty$. Then:

Theorem II. *For arbitrary initial conditions, under Assumptions I and II,*

$$(6) \qquad \lim_{t\to\infty} E\left\{\frac{\int_0^t \eta(x)\,dx}{t}\right\} = \frac{E\left\{\sum_{i=1}^{\nu-1} \omega_i\right\}}{E\{\gamma\}} \quad ,$$

whenever the limit on the RHS can be interpreted.

Proof:

Since $\{\eta(t) = 0\}$ is a recurrent event, it recurs infinitely often, with probability one; Assumption II makes the epochs which start a busy cycle the epochs of a (generalized) *renewal process*. $\int \eta(x)dx$ is then a *cumulative*, or *reward process* defined on the renewal process; use of the renewal theorem (see, for example, [2]), or of Tauberian theorems of transform calculus, leads to the well-known result that the limiting mean rate of accumulation is the mean accumulation per renewal, divided by the mean interval between renewals.

The only case in which the limit of (6) cannot be directly interpreted is when the RHS is of the form ∞/∞ .

Theorem II also holds, with probability one, for the time-average $(1/t)\int_0^t \eta(x)dx$ obtained from any realization.

Theorem III. For arbitrary initial conditions, under Assumptions I and II,

$$(7) \qquad \lim_{t\to\infty} \left\{\frac{\int_0^t \eta(x)\,dx}{t}\right\} \overset{a.s.}{=} \frac{E\left\{\sum_{i=0}^{\nu-1} \omega_i\right\}}{E\{\gamma\}}$$

whenever the limit on the RHS can be interpreted.

Proof:

Let $\gamma_1, \gamma_2, \ldots$ be the durations of successive busy cycles which begin at epochs $\Gamma_0 = 0$, $\Gamma_j = \sum_{i=1}^{j} \gamma_i$ $(j = 1,2,\ldots)$, and let $\varphi(t) = \sup\{j \mid \Gamma_j \le t\}$ be the number of busy cycles which start in $(0,t]$. Then, for t large enough so that $\varphi(t) > 0$,

$$\sum_{j=1}^{\varphi(t)} \int_{\Gamma_{j-1}}^{\Gamma_j} \eta(x)dx \le \int_0^t \eta(x)dx \le \sum_{j=1}^{\varphi(t)+1} \int_{\Gamma_{j-1}}^{\Gamma_j} \eta(x)dx$$

which can be rewritten as

$$\text{(8)} \quad \frac{\sum_{j=1}^{\varphi(t)} \int_{\Gamma_{j-1}}^{\Gamma_j} \eta(x)dx}{\varphi(t)} \cdot \frac{\varphi(t)}{t} \le \frac{\int_0^t \eta(x)dx}{t} < \frac{\sum_{j=1}^{\varphi(t)+1} \int_{\Gamma_{j-1}}^{\Gamma_j} \eta(x)dx}{\varphi(t)+1} \cdot \frac{\varphi(t)+1}{\varphi(t)} \cdot \frac{\varphi(t)}{t}$$

As $t\to\infty$, $\varphi(t)\to\infty$ with probability one since $\mathcal{E}$ is recurrent; indeed, $\varphi(t)/t \to 1/E\{\gamma\}$ with probability one (see, for example, [1] p. 51). Then, with probability one, the first term on both sides of (8) approaches $E\left\{\sum_{i=0}^{\nu-1} \omega_i\right\}$, by the strong law of large numbers, and the second term on the RHS of (8) approaches unity.

The final result requires:

Assumption III. For the distribution of Assumption II, the unit-average means

$$W = \frac{E\left\{\sum_{i=0}^{\nu-1} \omega_i\right\}}{E\{\nu\}} ; \qquad T = \frac{E\{\gamma\}}{E\{\nu\}} = \frac{E\left\{\sum_{i=1}^{\nu} \tau\right\}}{E\{\nu\}}$$

are both finite.

Then:

Theorem IV. For arbitrary initial conditions, under Assumptions I, II, and III,

$$\text{(10)} \qquad W = \lim_{n\to\infty} E\left\{\frac{1}{n} \sum_{i=0}^{n-1} \omega_i\right\} ; \quad T = \lim_{n\to\infty} E\left\{\frac{1}{n} \sum_{i=1}^{n} \tau_i\right\} ;$$

and

$$L \overset{\text{def}}{=} \lim_{t\to\infty} E\left\{\frac{\int_0^t \eta(x)dx}{t}\right\} = \frac{W}{T}$$

whenever T *is nonzero.*

Proof:

Consider a renewal process in discrete time, $n = 0,1,2, \ldots$, where a renewal occurs with the index of an arrival who starts a new busy period. If we consider $\sum_{i=0}^{n} \omega_i$ as the associated reward process, then the left-hand relation in (10) follows from the proof used in Theorem II, and the definition (9); similar remarks apply to $\sum_{i=1}^{n} \tau_i$. (11) follows from (6) and (9), whenever $E\{\nu\}$ is finite. If $E\{\nu\}$ infinite, a truncation argument will lead to (11) in the limit, providing T and W are well-defined.

The formula "$L = \lambda W$" then follows by defining the interarrival rate, λ, as Γ^{-1}.

DISCUSSION

As in [3], no specific assumptions are needed about independence of interarrival intervals, number of channels, service discipline, etc. The queueing system referred to may, in fact, be a portion of a larger system, such is the queueing units only, the members of one priorty class only, etc.

The assumption that the event $\mathcal{E} = \{\eta(t) = 0\}$ is recurrent is a most important one for our analysis, but is satisfied automatically by most assumptions of stationarity. Even if $\mathcal{E}$ is transient, Equation (4) still holds between any two occurences of the event. If it is known that some other state is recurrent, then the system may have a "steady-state" component which could be removed from the analysis.

Assumption II is almost always satisfied in simple queueing models, since the arrival mechanism is usually, although need not be, assumed independent of the service mechanism, and the service mechanism is usually "reset" by the first arrival after an idle period. This assumption can be weakened even if *not every* idle period resets the arrival and service mechanisms; we only need require that the process be reset *with probability one* after *some* idle period in the future, (and, hence, after infinitely many such idle periods).

We do not require that $E\{\iota\}$ be positive, although this, too, is usually satisfied in most queueing systems. For example, in the single-channel queue with identically distributed service times $\{\sigma_i\}$, $\beta = \sum_{i=0}^{\nu-1} \sigma_i$ in a given busy period, and a well-known resu . for W to be finite is that $E\{\beta\} < E\{\gamma\}$, or $E\{\sigma_i\} < T$; if the means are equal, and not both interarrival and service distributions are degenerate, then $E\{\iota\} = 0$, but $\mathcal{E}$ is still recurrent.

The assumption that W is finite means, of course, that $E\{\beta\}$ is finite if $E\{\nu\}$ is; nevertheless, the proof does not require that $E\{\nu\}$ be finite. If W is finite, but T is not, then (11) still holds with $L = 0$. Note that Assumption I may still hold (and L may even be finite) if *both* W and T are infinite, the event $\mathcal{E}$ then being *null-recurrent*. If both W and T are zero, then (6) may still provide the correct limit for L. (See Examples, Below).

The heart of the proof lies in Theorem I, which essentially states that, for a given curve $y(x)$ with well-defined beginning and end, the area can be calculated as $\int y\, dx$ or $\int x\, dy$. This result shows why we stress *time*-average inventories and *item*-average delays in operational models, but why the concepts of *virtual* delay, and inventory seen by an arrival, are not directly relevant.

EXAMPLES

Consider a bulk service mechanism which periodically, every H hours, sweeps out *all* waiting units. If units arrive in a Poisson stream, with mean spacing of

T hours, the unit-average delay is just $W = \frac{1}{2}H$. However, if n units arrive during one period, then by a well-known result on the conditional distribution of arrival epochs, the time-average number in the system (over one period, hence over all time) is just $\frac{1}{2}n$; unconditioning on the number of arrivals, we find $L = H/2T = W/T$. The average length of the idle period is T^{-1} ; even though the first new arrival after an idle period arrives during the middle of some period of length H , the service mechanism is reset in the sense that the distribution of wait for that start-up unit always has density $T^{-1}e^{t/T}(e^{H/T}-1)^{-1}$ for $0 \leq t < H$.

If the service mechanism sweeps out all *but one* unit whenever there are *two or more* units waiting, the assumptions are satisfied, even on a first-in-last-out basis, since the unit shoved aside merely has to wait $e^{h/T}-1$ periods, on the average, until no unit arrives, and he is the sole unit to be processed.

If the service mechanism *always* leaves this zeroth unit behind, Assumption 1 is not directly satisfied. However, the time-average number in the system is $L = 1 + (H/2T)$, and the value of W , defined by (10), is $\frac{1}{2}H + E\left\{\lim_{n\to\infty} \omega_0/n\right\} = \frac{1}{2}H + T^{-1}$, hence (11) is still correct! Or, we may note that the event $\{\eta(t) = 1\}$ is now recurrent, and can remove this unit from analysis by "ejecting" him for an arbitrarily small interval of idle time, and then bringing him back into the system.

If, in an M/G/1 queue, $E\{\sigma_i\} = T$, then we have both L and W , as well as $E\{\nu\}$, infinite; $\mathcal{E}$ is now a *null*-recurrent event. As another example where ν is infinite, let $\tau_i = (\frac{1}{2})^i (i=1,2, \ldots)$, and $\omega_i = \frac{5}{8} (\frac{1}{2})^i (i=0,1, \ldots)$, with the system reset at the limit point $t=1$. W and T in (10) are both zero, but $\sum_{i=0}^{\infty} \omega_i = \frac{5}{4}$, and $\gamma = 1$. Then, from (6), L should equal $\frac{5}{4}$; this is correct, since there is one unit in the system always, and two units during the intervals of length $1/8$, $1/16$, $1/32$, ... beginning at epochs $1/2$, $3/4$, $7/8$, ...

ACKNOWLEDGEMENTS

I would like to thank R. E. Barlow, J. D. C. Little, and R. W. Wolff for interesting discussions on this proof.

REFERENCES

[1] Barlow, R. E. and F. Proschan, MATHEMATICAL THEORY OF RELIABILITY, J. Wiley & Sons, New York, (1965).

[2] Jewell, W. S., "Markov-Renewal Programming, I & II," *Operations Research*, Vol. 11, No. 6, pp. 938-971, (November-December, 1963).

[3] Little, J. D. C., "A Proof for the Queueing Formula: $L = \lambda W$," *Operations Research*, Vol. 9, No. 3, pp. 383-387, (May-June, 1961).

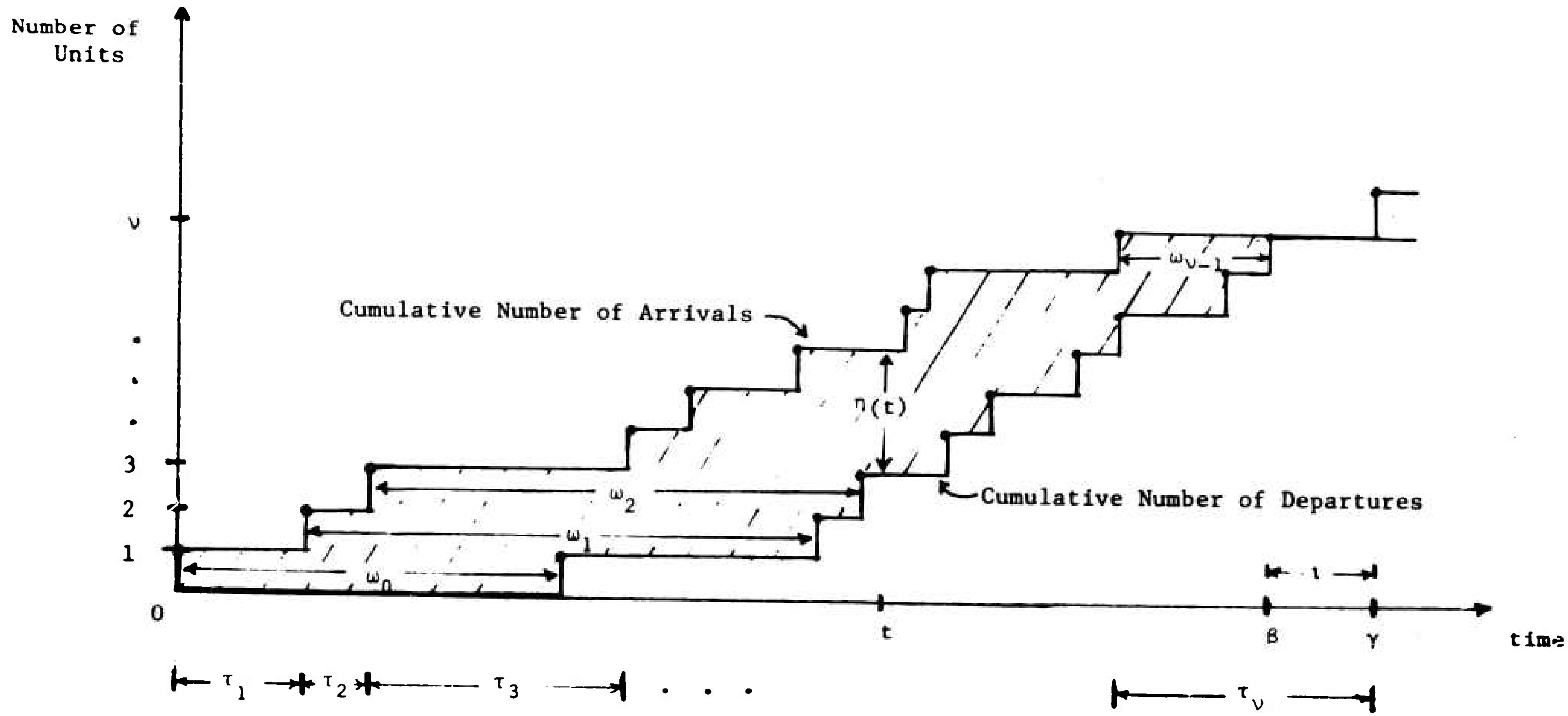

Figure 1. Typical Realization of Busy Cycle.

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author) University of California, Berkeley	2a REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b GROUP ---------

3 REPORT TITLE

"A Simple Proof of: L = λW"

4 DESCRIPTIVE NOTES (Type of report and inclusive dates)

RESEARCH REPORT December 1966

5 AUTHOR(S) (Last name, first name, initial)

Jewell, William, S.

6 REPORT DATE December 1966	7a TOTAL NO OF PAGES 9	7b NO OF REFS 3

8a CONTRACT OR GRANT NO Nonr-222(83) b PROJECT NO NR 047 033 c d Res. Proj. No: RR 003 07 01	9a ORIGINATOR'S REPORT NUMBER(S) O.R.C. 66-43 9b OTHER REPORT NO(S) (Any other numbers that may be assigned this report) none

10 AVAILABILITY LIMITATION NOTICES

DISTRIBUTION OF THIS DOCUMENT IS UNLIMITED.

11 SUPPLEMENTARY NOTES "Also supported by the Nat'l. Sci. Found. under Grant GP-4593, and the U.S. Army Research Office-Durham under Contract DA-31-124-ARO-D-331."	12 SPONSORING MILITARY ACTIVITY MATHEMATICAL SCIENCE DIVISION

13 ABSTRACT

A simple proof of the fundamental queueing formula L = λW is given which is based on renewal theory. The basic assumptions which are needed are: (1) the event {system is empty} is recurrent, and (2) the arrival and waiting-time mechanisms are reset by the next arrival after this event occurs.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
waiting-time						
queueing theory						
renewal-time						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year, or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through
____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through
____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through
____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, roles, and weights is optional.